JURA
Journal of Undergraduate Research in Alberta
UNIVERSITY OF CALGARY

PERSONALIZING MEDICINE: ANALYZING NEXT GENERATION SEQUENCING USING A SYMPTOMS BASED APPROACH ON PUBMED

Bruce Gao[1] & Matvei Privman [2]
[1]Hotchkiss Brain Institute, University of Calgary, [2]Independant Researcher
bmgao@ucalgary.ca

INTRODUCTION

Personalized medicine is the future of healthcare. Let us assume that there is a child with a rare genetic illness that results in cardiac arrest during teenage years. To best diagnose and treat this illness, doctors and scientists have to figure out which one of his 25,000 genes is defective. This modern version of finding a needle in a haystack is costly when patients are suffering.

Next Generation Sequencing (NGS) promises to accelerate this process and revolutionize medicine. Using advanced technologies to identify variants in whole genomes, NGS allows prediction and diagnosis of disease and personalized treatments to the individual [1]. Unfortunately, NGS identifies tens of thousands of variants: some real, some false positive and some false negative. Many variants can be excluded by comparing across genomes and rationalizing using different filter criteria. However, the list of potential variants involved in a given genetic disease remains several thousand long and we are back to looking for a needle in a haystack.

Overcoming this barrier would be a monumental advance in personalized medicine. We propose that the clues to finding the affective gene are hidden in massive mines of biomedical data online. Using this and the genetic information unique to each patient, we created an online platform that creates 'biograms' – personalized reports that identify disease causing genes based on the symptoms of a genetic disease.

METHODS

A physician or scientist navigates to www.biogram.co to access our platform. They enter two things: a list of gene variants from NGS that could be causing a genetic disease and a list of symptoms. Biogram then sends the information to a privately hosted Java server. The Java server is an observer that uses PubMed API (Entrez E-utilities) to populate a 2D search array with the number of hits for each symptom-gene combination. The Java server is robust, with error handling for corrupt input and network disconnections and able to handle over 1000 searches at once. The server then delivers the 'biogram' back to the physician or scientist who can use the information to guide diagnosis and treatment of a patient.

RESULTS

Part of a sample biogram is shown in figure 1. A biogram can be imported into Microsoft Excel for further analysis with heat maps and sorting. As seen below, the gene BRCA1 is implicated in diabetes and cancer. Furthermore, it seems like DRD5 is implicated in ADHD. There also seems to be an involvement of ABCG1 in diabetes. Indeed, a quick search on PubMed reveals that these connections have already been established in research. Biogram revealed these connections.

Variant	Hyperphagia	Diabetes	Hypoventilation	Apnea	Obesity	ADHD	Cancer
RETJ	0	0	0	0	0	0	0
ABCG1	1	66	0	0	23	0	58
ACOX2	0	1	0	0	0	0	4
ABLIM3	0	0	0	0	0	1	1
ABCC9	0	25	0	0	3	0	8
DRD5	0	1	0	0	2	64	7
ABCD1	0	1	0	0	1	4	15
ABHD10	0	0	0	0	0	0	0
BRCA1	0	75	0	0	43	0	10038

Figure 1. A biogram created at www.biogram.co using arbitrary symptoms and gene variants. The red circles indicate interesting gene hits for further analysis.

DISCUSSION AND CONCLUSIONS

In the introduction, a child with a rare genetic illness was discussed. In fact, this illness is real – it is called Rapid-onset Obesity with Hypothalamic dysfunction, Hypoventilation and Autonomic Dysregulation (ROHHAD). Biogram was used on a set of real Next Generation Sequencing data and a gene was found that was a variant in 5 of 11 individuals. Although not a complete solution, this highlights the promise of biograms and personalized medicine in the near future. Future goals include improving the user interface of Biogram and using advanced touch interfaces to sort, eliminate and highlight important genes.

REFERENCES

1. Newman & Black. *Genes.* **5**:1001-1017, 2014.
2. Sayers, E. *E-utilities quick start*, 2013.